#9

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/841,553A

DATE: 05/01/2003
TIME: 13:19:41

Input Set : A:\sequence listing.ST25.txt
Output Set: N:\CRF4\05012003\I841553A.raw

<110> APPLICANT: TAKAKURA, Hikaru
MORISHITA, Mio
YAMAMOTO, Katsuhiko
MITTA, Masanori
ASADA, Kiyozo
TSUNASAWA, Susumu
KATO, Ikunoshin
<120> TITLE OF INVENTION: ULTRATHERMOSTABLE PROTEASE GENES
<130> FILE REFERENCE: TAKAKURA=1A
<140> CURRENT APPLICATION NUMBER: 09/841,553A
<141> CURRENT FILING DATE: 2001-04-24
<150> PRIOR APPLICATION NUMBER: 08/894,818
<151> PRIOR FILING DATE: 1997-08-29
<150> PRIOR APPLICATION NUMBER: JP32385/1995
<151> PRIOR FILING DATE: 1995-12-12
<150> PRIOR APPLICATION NUMBER: JP96/03253
<151> PRIOR FILING DATE: 1996-11-07
<160> NUMBER OF SEQ ID NOS: 45
<170> SOFTWARE: PatentIn version 3.2
<210> SEQ ID NO: 1
<211> LENGTH: 659
<212> TYPE: PRT
<213> ORGANISM: Thermococcus celer
W--> <400> SEQUENCE: 1

```
Met Lys Arg Leu Gly Ala Val Val Leu Ala Leu Val Leu Val Gly Leu
1               5                  10                  15
Leu Ala Gly Thr Ala Leu Ala Ala Pro Val Lys Pro Val Val Arg Asn
            20                  25                  30
Asn Ala Val Gln Gln Lys Asn Tyr Gly Leu Leu Thr Pro Gly Leu Phe
        35                  40                  45
Lys Lys Val Gln Arg Met Asn Trp Asn Gln Glu Val Asp Thr Val Ile
    50                  55                  60
Met Phe Gly Ser Tyr Gly Asp Arg Asp Arg Ala Val Lys Val Leu Arg
65                  70                  75                  80
Leu Met Gly Ala Gln Val Lys Tyr Ser Tyr Lys Ile Ile Pro Ala Val
                85                  90                  95
Ala Val Lys Ile Lys Ala Arg Asp Leu Leu Leu Ile Ala Gly Met Ile
            100                 105                 110
Asp Thr Gly Tyr Phe Gly Asn Thr Arg Val Ser Gly Ile Lys Phe Ile
        115                 120                 125
Gln Glu Asp Tyr Lys Val Gln Val Asp Asp Ala Thr Ser Val Ser Gln
    130                 135                 140
Ile Gly Ala Asp Thr Val Trp Asn Ser Leu Gly Tyr Asp Gly Ser Gly
```

RAW SEQUENCE LISTING DATE: 05/01/2003
PATENT APPLICATION: US/09/841,553A TIME: 13:19:41

Input Set : A:\sequence listing.ST25.txt
Output Set: N:\CRF4\05012003\I841553A.raw

```
74 145                 150                 155                 160
77 Val Val Val Ala Ile Val Asp Thr Gly Ile Asp Ala Asn His Pro Asp
78                  165                 170                 175
81 Leu Lys Gly Lys Val Ile Gly Trp Tyr Asp Ala Val Asn Gly Arg Ser
82              180                 185                 190
85 Thr Pro Tyr Asp Asp Gln Gly His Gly Thr His Val Ala Gly Ile Val
86          195                 200                 205
89 Ala Gly Thr Gly Ser Val Asn Ser Gln Tyr Ile Gly Val Ala Pro Gly
90      210                 215                 220
93 Ala Lys Leu Val Gly Val Lys Val Leu Gly Ala Asp Gly Ser Gly Ser
94 225                 230                 235                 240
97 Val Ser Thr Ile Ile Ala Gly Val Asp Trp Val Val Gln Asn Lys Asp
98                  245                 250                 255
101 Lys Tyr Gly Ile Arg Val Ile Asn Leu Ser Leu Gly Ser Ser Gln Ser
102             260                 265                 270
105 Ser Asp Gly Thr Asp Ser Leu Ser Gln Ala Val Asn Asn Ala Trp Asp
106         275                 280                 285
109 Ala Gly Ile Val Val Cys Val Ala Ala Gly Asn Ser Gly Pro Asn Thr
110     290                 295                 300
113 Tyr Thr Val Gly Ser Pro Ala Ala Ala Ser Lys Val Ile Thr Val Gly
114 305                 310                 315                 320
117 Ala Val Asp Ser Asn Asp Asn Ile Ala Ser Phe Ser Ser Arg Gly Pro
118                 325                 330                 335
121 Thr Ala Asp Gly Arg Leu Lys Pro Glu Val Val Ala Pro Gly Val Asp
122             340                 345                 350
125 Ile Ile Ala Pro Arg Ala Ser Gly Thr Ser Met Gly Thr Pro Ile Asn
126         355                 360                 365
129 Asp Tyr Tyr Thr Lys Ala Ser Gly Thr Ser Met Ala Thr Pro His Val
130     370                 375                 380
133 Ser Gly Val Gly Ala Leu Ile Leu Gln Ala His Pro Ser Trp Thr Pro
134 385                 390                 395                 400
137 Asp Lys Val Lys Thr Ala Leu Ile Glu Thr Ala Asp Ile Val Ala Pro
138                 405                 410                 415
141 Lys Glu Ile Ala Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Val Tyr
142             420                 425                 430
145 Lys Ala Ile Lys Tyr Asp Asp Tyr Ala Lys Leu Thr Phe Thr Gly Ser
146         435                 440                 445
149 Val Ala Asp Lys Gly Ser Ala Thr His Thr Phe Asp Val Ser Gly Ala
150     450                 455                 460
153 Thr Phe Val Thr Ala Thr Leu Tyr Trp Asp Thr Gly Ser Ser Asp Ile
154 465                 470                 475                 480
157 Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Glu Val Asp Tyr Ser Tyr
158                 485                 490                 495
161 Thr Ala Tyr Tyr Gly Phe Glu Lys Val Gly Tyr Tyr Asn Pro Thr Ala
162             500                 505                 510
165 Gly Thr Trp Thr Val Lys Val Val Ser Tyr Lys Gly Ala Ala Asn Tyr
166         515                 520                 525
169 Gln Val Asp Val Val Ser Asp Gly Ser Leu Ser Gln Ser Gly Gly Gly
170     530                 535                 540
```

RAW SEQUENCE LISTING DATE: 05/01/2003
PATENT APPLICATION: US/09/841,553A TIME: 13:19:41

Input Set : A:\sequence listing.ST25.txt
Output Set: N:\CRF4\05012003\I841553A.raw

```
173 Asn Pro Asn Pro Asn Pro Asn Pro Asn Pro Thr Pro Thr Thr Asp Thr
174 545                 550                 555                 560
177 Gln Thr Phe Thr Gly Ser Val Asn Asp Tyr Trp Asp Thr Ser Asp Thr
178                 565                 570                 575
181 Phe Thr Met Asn Val Asn Ser Gly Ala Thr Lys Ile Thr Gly Asp Leu
182             580                 585                 590
185 Thr Phe Asp Thr Ser Tyr Asn Asp Leu Asp Leu Tyr Leu Tyr Asp Pro
186         595                 600                 605
189 Asn Gly Asn Leu Val Asp Arg Ser Thr Ser Ser Asn Ser Tyr Glu His
190     610                 615                 620
193 Val Glu Tyr Ala Asn Pro Ala Pro Gly Thr Trp Thr Phe Leu Val Tyr
194 625                 630                 635                 640
197 Ala Tyr Ser Thr Tyr Gly Trp Ala Asp Tyr Gln Leu Lys Ala Val Val
198                 645                 650                 655
201 Tyr Tyr Gly
205 <210> SEQ ID NO: 2
206 <211> LENGTH: 1977
207 <212> TYPE: DNA
208 <213> ORGANISM: Thermococcus celer
210 <400> SEQUENCE: 2
211 atgaagaggt taggtgctgt ggtgctggca ctggtgctcg tgggtcttct ggccggaacg     60
213 gcccttgcgg cacccgtaaa accggttgtc aggaacaacg cggttcagca gaagaactac    120
215 ggactgctga ccccgggact gttcaagaaa gtccagagga tgaactggaa ccaggaagtg    180
217 gacaccgtca taatgttcgg gagctacgga gacagggaca gggcggttaa ggtactgagg    240
219 ctcatgggcg cccaggtcaa gtactcctac aagataatcc ctgctgtcgc ggttaaaata    300
221 aaggccaggg accttctgct gatcgcgggc atgatagaca cgggttactt cggtaacaca    360
223 agggtctcgg gcataaagtt catacaggag gattacaagg ttcaggttga cgacgccact    420
225 tccgtctccc agataggggc cgataccgtc tggaactccc tcggctacga cggaagcggt    480
227 gtggtggttg ccatcgtcga tacgggtata gacgcgaacc accccgatct gaagggcaag    540
229 gtcataggct ggtacgacgc cgtcaacggc aggtcgaccc cctacgatga ccagggacac    600
231 ggaacccacg ttgcgggtat cgttgccgga accggcagcg ttaactccca gtacataggc    660
233 gtcgcccccg gcgcgaagct cgtcggcgtc aaggttctcg gtgccgacgg ttcgggaagc    720
235 gtctccacca tcatcgcggg tgttgactgg gtcgtccaga acaaggacaa gtacgggata    780
237 agggtcatca acctctccct cggctcctcc cagagctccg acggaaccga ctccctcagt    840
239 caggccgtca acaacgcctg ggacgccggt atagtagtct gcgtcgccgc cggcaacagc    900
241 gggccgaaca cctacaccgt cggctcaccc gccgccgcga gcaaggtcat aaccgtcggt    960
243 gcagttgaca gcaacgacaa catcgccagc ttctccagca ggggaccgac cgcggacgga   1020
245 aggctcaagc cggaagtcgt cgcccccggc gttgacatca tagccccgcg cgccagcgga   1080
247 accagcatgg gcaccccgat aaacgactac tacaccaagg cctctggaac cagcatggcc   1140
249 accccgcacg tttcgggcgt tggcgcgctc atcctccagg cccacccgag ctggaccccg   1200
251 gacaaggtga agaccgccct catcgagacc gccgacatag tcgcccccaa ggagatagcg   1260
253 gacatcgcct acggtgcggg tagggtgaac gtctacaagg ccatcaagta cgacgactac   1320
255 gccaagctca ccttcaccgg ctccgtcgcc gacaagggaa gcgccaccca caccttcgac   1380
257 gtcagcggcg ccaccttcgt gaccgccacc ctctactggg acacgggctc gagcgacatc   1440
259 gacctctacc tctacgaccc caacgggaac gaggttgact actcctacac cgcctactac   1500
261 ggcttcgaga aggtcggcta ctacaacccg accgccggaa cctggacggt caaggtcgtc   1560
263 agctacaagg gcgcggcgaa ctaccaggtc gacgtcgtca gcgacgggag cctcagccag   1620
265 tccggcggcg gcaacccgaa tccaaacccc aacccgaacc caaccccgac caccgacacc   1680
267 cagaccttca ccggttccgt taacgactac tgggacacca gcgacacctt caccatgaac   1740
```

RAW SEQUENCE LISTING DATE: 05/01/2003
PATENT APPLICATION: US/09/841,553A TIME: 13:19:41

Input Set : A:\sequence listing.ST25.txt
Output Set: N:\CRF4\05012003\I841553A.raw

```
269 gtcaacagcg gtgccaccaa gataaccggt gacctgacct tcgatacttc ctacaacgac   1800
271 ctcgacctct acctctacga ccccaacggc aacctcgttg acaggtccac gtcgagcaac   1860
273 agctacgagc acgtcgagta cgccaacccc gccccgggaa cctggacgtt cctcgtctac   1920
275 gcctacagca cctacggctg ggcggactac cagctcaagg ccgtcgtcta ctacggg      1977
278 <210> SEQ ID NO: 3
279 <211> LENGTH: 522
280 <212> TYPE: PRT
281 <213> ORGANISM: Pyrococcus furiosus
284 <220> FEATURE:
285 <221> NAME/KEY: misc_feature
286 <222> LOCATION: (428)..(428)
287 <223> OTHER INFORMATION: Xaa is Gly or Val
289 <400> SEQUENCE: 3
291 Ala Glu Leu Glu Gly Leu Asp Glu Ser Ala Ala Gln Val Met Ala Thr
292 1               5                  10                  15
295 Tyr Val Trp Asn Leu Gly Tyr Asp Gly Ser Gly Ile Thr Ile Gly Ile
296             20                  25                  30
299 Ile Asp Thr Gly Ile Asp Ala Ser His Pro Asp Leu Gln Gly Lys Val
300         35                  40                  45
303 Ile Gly Trp Val Asp Phe Val Asn Gly Arg Ser Tyr Pro Tyr Asp Asp
304     50                  55                  60
307 His Gly His Gly Thr His Val Ala Ser Ile Ala Ala Gly Thr Gly Ala
308 65                  70                  75                  80
311 Ala Ser Asn Gly Lys Tyr Lys Gly Met Ala Pro Gly Ala Lys Leu Ala
312                 85                  90                  95
315 Gly Ile Lys Val Leu Gly Ala Asp Gly Ser Gly Ser Ile Ser Thr Ile
316             100                 105                 110
319 Ile Lys Gly Val Glu Trp Ala Val Asp Asn Lys Asp Lys Tyr Gly Ile
320         115                 120                 125
323 Lys Val Ile Asn Leu Ser Leu Gly Ser Ser Gln Ser Ser Asp Gly Thr
324     130                 135                 140
327 Asp Ala Leu Ser Gln Ala Val Asn Ala Ala Trp Asp Ala Gly Leu Val
328 145                 150                 155                 160
331 Val Val Val Ala Ala Gly Asn Ser Gly Pro Asn Lys Tyr Thr Ile Gly
332                 165                 170                 175
335 Ser Pro Ala Ala Ala Ser Lys Val Ile Thr Val Gly Ala Val Asp Lys
336             180                 185                 190
339 Tyr Asp Val Ile Thr Ser Phe Ser Ser Arg Gly Pro Thr Ala Asp Gly
340         195                 200                 205
343 Arg Leu Lys Pro Glu Val Val Ala Pro Gly Asn Trp Ile Ile Ala Ala
344     210                 215                 220
347 Arg Ala Ser Gly Thr Ser Met Gly Gln Pro Ile Asn Asp Tyr Tyr Thr
348 225                 230                 235                 240
351 Ala Ala Pro Gly Thr Ser Met Ala Thr Pro His Val Ala Gly Ile Ala
352                 245                 250                 255
355 Ala Leu Leu Leu Gln Ala His Pro Ser Trp Thr Pro Asp Lys Val Lys
356             260                 265                 270
359 Thr Ala Leu Ile Glu Thr Ala Asp Ile Val Lys Pro Asp Glu Ile Ala
360         275                 280                 285
```

RAW SEQUENCE LISTING DATE: 05/01/2003
PATENT APPLICATION: US/09/841,553A TIME: 13:19:41

Input Set : A:\sequence listing.ST25.txt
Output Set: N:\CRF4\05012003\I841553A.raw

```
363 Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Ala Tyr Lys Ala Ile Asn
364     290                 295                 300
367 Tyr Asp Asn Tyr Ala Lys Leu Val Phe Thr Gly Tyr Val Ala Asn Lys
368 305                 310                 315                 320
371 Gly Ser Gln Thr His Gln Phe Val Ile Ser Gly Ala Ser Phe Val Thr
372                 325                 330                 335
375 Ala Thr Leu Tyr Trp Asp Asn Ala Asn Ser Asp Leu Asp Leu Tyr Leu
376             340                 345                 350
379 Tyr Asp Pro Asn Gly Asn Gln Val Asp Tyr Ser Tyr Thr Ala Tyr Tyr
380         355                 360                 365
383 Gly Phe Glu Lys Val Gly Tyr Tyr Asn Pro Thr Asp Gly Thr Trp Thr
384     370                 375                 380
387 Ile Lys Val Val Ser Tyr Ser Gly Ser Ala Asn Tyr Gln Val Asp Val
388 385                 390                 395                 400
391 Val Ser Asp Gly Ser Leu Ser Gln Pro Gly Ser Ser Pro Ser Pro Gln
392                 405                 410                 415
W--> 395 Pro Glu Pro Thr Val Asp Ala Lys Thr Phe Gln Xaa Ser Asp His Tyr
396             420                 425                 430
399 Tyr Tyr Asp Arg Ser Asp Thr Phe Thr Met Thr Val Asn Ser Gly Ala
400         435                 440                 445
403 Thr Lys Ile Thr Gly Asp Leu Val Phe Asp Thr Ser Tyr His Asp Leu
404     450                 455                 460
407 Asp Leu Tyr Leu Tyr Asp Pro Asn Gln Lys Leu Val Asp Arg Ser Glu
408 465                 470                 475                 480
411 Ser Pro Asn Ser Tyr Glu His Val Glu Tyr Leu Thr Pro Ala Pro Gly
412                 485                 490                 495
415 Thr Trp Tyr Phe Leu Val Tyr Ala Tyr Tyr Thr Tyr Gly Trp Ala Tyr
416             500                 505                 510
419 Tyr Glu Leu Thr Ala Lys Val Tyr Tyr Gly
420         515                 520
423 <210> SEQ ID NO: 4
424 <211> LENGTH: 1566
425 <212> TYPE: DNA
426 <213> ORGANISM: Pyrococcus furiosus
429 <220> FEATURE:
430 <221> NAME/KEY: misc_feature
431 <222> LOCATION: (1283)..(1283)
432 <223> OTHER INFORMATION: n is G or T
434 <400> SEQUENCE: 4
435 gcagaattag aaggactgga tgagtctgca gctcaagtta tggcaactta cgtttggaac     60
437 ttgggatatg atggttctgg aatcacaata ggaataattg acactggaat tgacgcttct    120
439 catccagatc tccaaggaaa agtaattggg tgggtagatt ttgtcaatgg taggagttat    180
441 ccatacgatg accatggaca tggaactcat gtagcttcaa tagcagctgg tactggagca    240
443 gcaagtaatg gcaagtacaa gggaatggct ccaggagcta agctggcggg aattaaggtt    300
445 ctaggtgccg atggttctgg aagcatatct actataatta agggagttga gtgggccgtt    360
447 gataacaaag ataagtacgg aattaaggtc attaatcttt ctcttggttc aagccagagc    420
449 tcagatggta ctgacgctct aagtcaggct gttaatgcag cgtgggatgc tggattagtt    480
451 gttgtggttg ccgctggaaa cagtggacct aacaagtata caatcggttc tccagcagct    540
453 gcaagcaaag ttattacagt tggagccgtt gacaagtatg atgttataac aagcttctca    600
```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 05/01/2003
PATENT APPLICATION: US/09/841,553A TIME: 13:19:42

Input Set : A:\sequence listing.ST25.txt
Output Set: N:\CRF4\05012003\I841553A.raw

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:3; Xaa Pos. 428
Seq#:4; N Pos. 1283
Seq#:13; N Pos. 127,130,136

Invalid <213> Response:
Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:43,44